



(12) **EUROPEAN PATENT APPLICATION**  
 published in accordance with Art. 153(4) EPC

(43) Date of publication:  
**10.05.2017 Bulletin 2017/19**

(51) Int Cl.:  
**G01N 27/62 (2006.01)**

(21) Application number: **15815598.6**

(86) International application number:  
**PCT/JP2015/050622**

(22) Date of filing: **13.01.2015**

(87) International publication number:  
**WO 2016/002233 (07.01.2016 Gazette 2016/01)**

(84) Designated Contracting States:  
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**  
 Designated Extension States:  
**BA ME**

(72) Inventors:  
 • **KAGEYAMA, Tetsuya**  
**Kyoto-shi, Kyoto 604-8511 (JP)**  
 • **U MEMURA, Yoshikatsu**  
**Kyoto-shi, Kyoto 604-8511 (JP)**

(30) Priority: **03.07.2014 PCT/JP2014/067813**

(74) Representative: **Webster, Jeremy Mark et al**  
**Mewburn Ellis LLP**  
**City Tower**  
**40 Basinghall Street**  
**London EC2V 5DE (GB)**

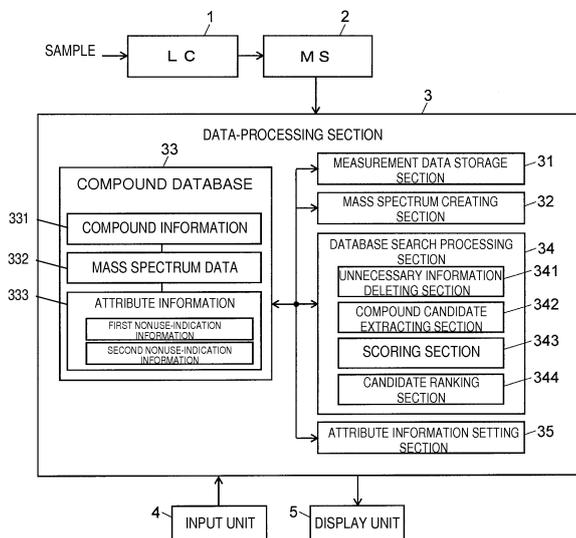
(71) Applicant: **Shimadzu Corporation**  
**Kyoto 604-8511 (JP)**

(54) **MASS-SPECTROMETRY-DATA PROCESSING DEVICE**

(57) Nonuse-indication information can be set for each peak on a mass spectrum collected in a compound database (33), as attribute information, the nonuse-indication information allowing the selection of whether to be used in a database search. For example, nonuse-indication information is set in advance to a noise peak mixed in actual measurement, an impurity-originated peak, and the like. In identifying a compound, when nonuse-indication information is read from the database (33) for a database search together with a mass spectrum, an unnecessary

information deleting section (341) transmits a mass spectrum from which a peak set with the nonuse-indication information is deleted, to a compound candidate extracting section (342) and a scoring section (343). Therefore, a peak set with the nonuse-indication information is ignored in, for example, calculating the score of a compound candidate, which allows a score with a high accuracy to be calculated, resulting in an improved identification accuracy.

Fig. 1



## Description

### TECHNICAL FIELD

**[0001]** The present invention relates to a data-processing system that analyzes data collected with a mass spectrometer, more specifically, relates to a data-processing system that identifies an unknown compound in a sample or estimates the structure of the unknown compound, using a database in which mass spectra corresponding to a large number of known compounds are collected.

### BACKGROUND ART

**[0002]** Methods for identifying an unknown compound in a sample using a gas chromatograph mass spectrometer (GC-MS) or a liquid chromatograph mass spectrometer (LC-MS) include a well-known method involving a database search using a database (may be called a library) in which mass spectra (including MS<sup>n</sup> spectra, where n is an integer of two or more) corresponding to a large number of known compounds are collected. Databases in which such mass spectra are collected range from general-purpose databases being exhaustive collection of the mass spectra of general compounds, such as the NIST database compiled by the National Institute of Standards and Technology (NIST) (US), and the Wiley database compiled by John Wiley & Sons, Inc., a publisher, to specified databases of compounds in specific fields or compounds for specific purposes, such as databases for agricultural chemical, medicine, and metabolite (see Patent Literature 1, etc.).

**[0003]** Such databases of mass spectra are created in general based on data obtained by actually measuring standard preparations of target compounds using a measurement machine. Usually, in collecting data, a mass spectrum includes unnecessary elements such as noise due to various factors. For example, in an LC-MS, ions originating from impurities included in a mobile phase used in an LC may appear on a mass spectrum in the form of the unnecessary elements. In addition, ions originating from impurities separated out from a column may also appear on a mass spectrum in the form of unnecessary elements. As described above, it is possible that unnecessary elements are included in a mass spectrum, but it is not desirable in terms of the reliability of analysis to edit the obtained mass spectrum before storing it to a database. Therefore, even when including unnecessary elements, such a mass spectrum is usually collected in a database as it is.

**[0004]** While there are a variety of algorithms for searching database based on mass spectra, algorithms commonly include two steps: extracting, from a database, a plurality of compounds the mass spectra of which have similar spectral patterns to that of an unknown compound to some extent, as compound candidates; and then calculating scores of strict degrees of match of the compound candidates in spectral pattern so that the com-

ound candidates are presented to a user in a descending order of the scores. In performing such data processing, unnecessary elements included in a mass spectrum in the database can lead to the presentation of a false positive or a false negative, which may decrease the accuracy of a search.

**[0005]** On mass spectra of similar compounds having a common main skeleton, a common spectral pattern originating from the common main skeleton appears, which makes the mass spectra significantly similar to one another. In the case of a database in which a large number of such similar compounds are collected, for an unknown compound being one of such similar compounds, a large number of compounds similar to one another in mass spectrum are extracted as compound candidates. The spectral patterns of mass spectra of a plurality of compound candidates extracted in such a manner share a lot of common portions, and thus calculated scores hardly yield significant differences from one another, which makes it difficult to assess which of the compound candidates is a correct compound even when the scores are compared. In addition, this reduces the possibility that the correct compound is highly ranked. Furthermore, in identifying a compound using mass spectra, an analysis operator often determines the final result by visually confirming the match of a mass spectrum, but if the number of extracted compound candidates is too large, the visual confirmation operation poses a heavy workload on the operator and an operation error such as overlooking easily occurs.

### CITATION LIST

#### PATENT LITERATURE

**[0006]** [Patent Literature 1] JP 2005-83952 A

### SUMMARY OF INVENTION

#### TECHNICAL PROBLEM

**[0007]** The present invention has been developed to solve the previously-described problem. Its objective is to provide a mass spectrometry data-processing system capable of improving the accuracy of identifying a compound or estimating a structure by searching a database even when an unnecessary element such as noise is included in a mass spectrum collected in a database.

**[0008]** Another objective of the present invention is to provide a mass spectrometry data-processing system capable of, even when a large number of similar compounds having a common main skeleton are collected in a database and one of the compounds is a target compound, eliminating the influence of similar compounds in a database search, so as to identify a target compound with exactitude.

## SOLUTION TO PROBLEM

**[0009]** A first specific form of the present invention developed for solving the previously-described problem is a mass spectrometry data-processing system that estimates the substance or the structure of an unknown compound by checking a mass spectrum obtained by subjecting the unknown compound to mass spectrometry against a database in which the mass spectra of known compounds are collected, the mass spectrometry data-processing system including:

- a) a database in which mass spectra of known compounds are collected, and nonuse-indication information indicating that it is not used in a step in a later-described database search can be set as attribute information in association with at least some of ion peaks appearing on a mass spectrum; and
- b) a search executing section for performing a database search on a mass spectrum obtained for the unknown compound using the database through the use of a mass spectrum from which an ion peak with the attribute information of the nonuse-indication information is excluded when performing the database search.

**[0010]** The term "mass spectrometry" used herein includes MS/MS analysis including fragmentation operation on ions and MS<sup>n</sup> analysis, where n is three or more, and the term "mass spectrum" used herein includes an MS/MS spectrum and an MS<sup>n</sup> spectrum.

**[0011]** In the mass spectrometry data-processing system according to the first specific form of the present invention, the mass spectra collected in the database are normally line spectra having been subjected to centroid processing. For example, in an ion source employing electron ionization, a molecular ion is easily fragmented to generate various fragment ions in ionization, and thus a peak originating from a molecular ion as well as various fragment ion peaks appear on a mass spectrum. In addition, since many elements have natural isotopes, a molecular ion peak or a fragment ion peak is accompanied with a monoisotopic mass peak as well as one or more isotope peaks. That is, on a mass spectrum, ion peaks appear at a plurality of different positions on the mass-to-charge ratio m/z axis. The database in the mass spectrometry data-processing system according to the first specific form of the present invention is configured to allow, for example, nonuse-indication information to be set for each ion peak, in other words, for each mass-to-charge ratio on a mass spectrum, as attribute information.

**[0012]** Nonuse-indication information may be set when the mass spectrum is collected in the database or may be set after the mass spectrum is collected in the database. In addition, nonuse-indication information may be newly set to a mass spectrum having been already collected in the database, or nonuse-indication information

having been set to a mass spectrum may be changed. In this case, security measures are desirably taken so that the nonuse-indication information will not be changed due to an error or intentionally.

**[0013]** For example, at the time of collecting, in a database, a mass spectrum obtained by performing actual measurement on a known compound, when there is an ion peak clearly estimated that the ion peak should not be generated from the compound, that is, clearly estimated to be noise, setting nonuse-indication information to the ion peak may suffice.

**[0014]** When an unknown compound in a sample is to be identified, and a mass spectrum obtained by subjecting the unknown compound to mass spectrometry is given, the search executing section performs a database search on the mass spectrum. At this point, attribute information is checked for each mass spectrum collected in the database, and when there is an ion peak attached with nonuse-indication information on a mass spectrum, the mass spectrum from which the ion peak is excluded is to be subjected to a comparison with the mass spectrum of the unknown compound. With this configuration, as previously described, when a noise peak is attached with nonuse-indication information, the noise peak is not involved in the comparison of spectral patterns or the like, which avoids a decrease in the accuracy of searches due to the presence of a noise peak.

**[0015]** As a preferable specific form of embodiment of the mass spectrometry data-processing system according to the present invention, the search executing section preferably includes: a candidate extracting section for comparing the patterns of mass spectra to extract a plurality of compounds to be candidates; and a ranking section for ranking the plurality of extracted candidate compounds based on the similarities of mass spectra,

the nonuse-indication information preferably includes: first nonuse-indication information indicating not being used in both of extracting the compounds by the candidate extracting section and ranking the compound candidates by the ranking section; and second nonuse-indication information indicating not being used only in ranking the compound candidates by the ranking section, and the candidate extracting section preferably extracts a plurality of compound candidates using mass spectra from which an ion peak attached with the first nonuse-indication information is excluded, and the ranking section preferably ranks compound candidates using mass spectrum from which an ion peak attached with at least one of the first nonuse-indication information and the second nonuse-indication information is excluded.

**[0016]** In this configuration, the processing of a database search performed by the search executing section includes two steps: a step of extracting a plurality of compound candidates by the comparison of spectral patterns; and a step of calculating scores indicating the similarities of the compound candidates, and ranking the compound candidates, for example, in accordance with the scores.

To distinguish an ion peak not to be used in both steps and an ion peak to be used in the step of extracting the compound candidates but not to be used in the step of ranking, two kinds of nonuse-indication information are assigned. Therefore, in the database search process by the search executing section, any specific ion peak appearing on a mass spectrum in a database can be ignored, or a database search can be performed in such a manner that a specific ion peak appearing on a mass spectrum in the database is taken into account in the step of extracting the plurality of compound candidate, whereas the same ion peak is ignored in the step of ranking.

**[0017]** For example, in the case where a large number of similar compounds having a common main skeleton are present in a database, and one of such compounds is a target compound being an identification object, an ion peak corresponding to the main skeleton may be attached with second nonuse-indication information. In this case, in the step of extracting a plurality of compound candidates, the ion peak corresponding to the main skeleton is taken into account, and thus a large number of similar compounds having the same main skeleton as the main skeleton of the target compound are extracted as the compound candidates. Meanwhile, when the score of each of the compound candidate extracted in such a manner is calculated, an ion peak corresponding to the same main skeleton is ignored, and thus the similarity and the dissimilarity of an ion peak characteristic of each compound other than the main skeleton is significantly reflected in the score.

**[0018]** As a result, even when a large number of similar compounds are present, a correct compound out of the similar compounds is likely to be given a high score, which allows the correct compound to be obtained as an identification result accurately. In addition, compound candidates can be narrowed accurately in the step of ranking, which can reduce a workload on the operator when the operator performs the visual confirmation of a match of a mass spectrum. Furthermore, the occurrence of a confirmation error or the like by the operator can be reduced.

**[0019]** A second specific form of the present invention developed for solving the previously described problem is a mass spectrometry data-processing system that estimates an unknown compound with a database search using a database in which mass spectra of known compounds are collected, the mass spectrometry data-processing system including:

- a) a database in which mass spectra are collected, and priority indication information indicating that it is used preferentially in a database search can be set as attribute information in association with at least some of ion peaks appearing on a mass spectrum; and
- b) a search processing section for performing a database search by comparing a mass spectrum of an unknown compound with mass spectra in the data-

base to extract compound candidates that is supposed to be an unknown compound of interest, the search processing section extracting the compound candidates on one of the conditions that an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound, or creating a search result that allows, of compound candidates extracted through a database search not using the priority indication information, a compound candidate of which an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound to be distinguished from the other compound candidates.

**[0020]** In the mass spectrometry data-processing system according to the second specific form of the present invention, as with the nonuse-indication information in the first specific form, for example, priority indication information can be set for each ion peak, that is, for each mass-to-charge ratio on a mass spectrum, as one of the kinds of attribute information.

**[0021]** When an unknown compound in a sample is to be identified, and a mass spectrum obtained by subjecting the unknown compound to mass spectrometry is given, the search executing section performs, for example, a normal database search on the mass spectrum to extract a plurality of compound candidates having high similarities in spectral pattern. Subsequently, the search executing section performs the confirmation of whether or not there is any ion peak set with priority indication information on mass spectra corresponding to the plurality of extracted compound candidates, and excludes a compound candidate of which the ion peak set with priority indication information is absent on the mass spectrum of the unknown compound. Alternatively, instead of excluding, the search executing section creates and displays a search result so that, of the plurality of compound candidate extracted through the database search, a compound candidate of which an ion peak set with priority indication information is present on the mass spectrum of the unknown compound can be distinguished from the other compound candidates.

**[0022]** Therefore, in the case where there are a large number of similar but different compounds having a common main skeleton, setting priority indication information to not an ion peak corresponding to such a main skeleton but to other ion peaks characteristic of partial structures of individual compounds allows compound candidates having a partial structure present in an unknown compound, that is, being highly likely to be correct, to be included in final compound candidates or to be presented to the operator so as to be distinguished from the other compound candidates.

**[0023]** In the mass spectrometry data-processing system according to the second specific form of the present invention, more preferably, the database may be configured to allow threshold information to be set, as attribute information, to an ion peak

set with the priority indication information, in addition to the priority indication information, the threshold information being used for assessing the intensity ratio of the signal intensity of the ion peak to the signal intensity of a specific peak on a mass spectrum in which the ion peak is present.

**[0024]** Here, the "specific peak" may be set to a peak at a specific mass-to-charge ratio on a mass spectrum, or may be set to a peak showing a maximum intensity irrespective of the mass-to-charge ratio.

**[0025]** Then, in this configuration, the search processing section may be configured to extract compound candidates on one of the conditions that an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound, and, on the mass spectrum of the unknown compound, the intensity ratio of the signal intensity of the ion peak to the signal intensity of the specific peak is equal to or greater than the threshold information set to the ion peak together with the priority indication information.

**[0026]** With this configuration, even when an ion peak set with priority indication information is observed on the mass spectrum of the unknown compound, a compound candidate including an ion peak the signal intensity of which is assumed to be insignificant is substantially excluded from the candidates of the unknown compound. This allows the compound candidates to be accurately narrowed, which in turn allows the operator to specify a correct compound easily.

#### ADVANTAGEOUS EFFECTS OF INVENTION

**[0027]** With the mass spectrometry data-processing system according to the first and second specific forms of the present invention, it is possible to improve, while using mass spectra obtained through actual measurement in a database, the accuracy of the identification and the structure analysis of a compound through a database search using this database.

#### BRIEF DESCRIPTION OF DRAWINGS

##### **[0028]**

Fig. 1 is a schematic configuration diagram of a first embodiment of an LC-MS system including a mass spectrometry data-processing system according to the present invention.

Fig. 2A, Fig. 2B, and Fig. 2C are diagrams illustrating examples of a compound identification process in the LC-MS system in the first embodiment.

Fig. 3 is a schematic configuration diagram of a second embodiment of an LC-MS system including a mass spectrometry data-processing system according to the present invention.

Fig. 4 is a diagram illustrating an example of a compound identification process in the LC-MS system in the second embodiment.

#### DESCRIPTION OF EMBODIMENTS

**[0029]** A first embodiment of an LC-MS system using a mass spectrometry data-processing system according to the present invention is described with reference to the attached drawings. Fig. 1 is a schematic configuration diagram of the LC-MS system of the present embodiment, and Fig. 2A, Fig. 2B, and Fig. 2C are diagrams illustrating examples of a compound identification process in the LC-MS system in the present embodiment.

**[0030]** The LC-MS system in the present embodiment includes a liquid chromatograph (LC) 1, a mass spectrometry section (MS) 2, and a data-processing section 3.

**[0031]** Although not being illustrated, the liquid chromatograph 1 includes, as basic components for a liquid chromatograph, a mobile phase container in which a mobile phase is stored, a pump for supplying the mobile phase at a certain flow velocity, an injector for injecting a liquid sample in a predetermined amount into the supplied mobile phase, a column for separating compounds contained in the liquid sample, and the like. The liquid chromatograph 1 transports the sample containing temporally separated compounds to the mass spectrometry section 2.

**[0032]** Although not being illustrated as with the liquid chromatograph 1, the mass spectrometry section 2 includes an atmospheric pressure ion source for ionizing compounds in a sample, an ion guide for transporting generated ions, a quadrupole mass filter for separating the ions in proportion to their mass-to-charge ratios  $m/z$ , an ion detector for detecting the separated ions, and the like. By repeating scan measurement over a predetermined mass-to-charge-ratio range using the quadrupole mass filter, the mass spectrometry section 2 can repeatedly collect mass spectrum data within the mass-to-charge-ratio range.

**[0033]** The data-processing section 3 includes, as functional blocks, a measurement data storage section 31, a mass spectrum creating section 32, a compound database 33, a database search processing section 34, and an attribute information setting section 35, and the database search processing section 34 includes an unnecessary information deleting section 341, a compound candidate extracting section 342, a scoring section 343, and ranking section 344. To the data-processing section 3, an input unit 4 operated by an analysis operator for setting a searching condition in a database search and a display unit 5 for displaying the searching condition, a search result, and the like are connected.

**[0034]** Part or most of the functions of the data-processing section 3 can be implemented by executing predetermined data processing programs on a personal computer.

**[0035]** In the compound database 33, mass spectrum data 332 on a large number of compounds is registered being associated with compound information 331 containing a compound name, a structural formula, a molecular weight, and the like. This is the same as conventional

systems. Here, attribute information 333 on the mass spectrum data 332 can be further registered. The attribute information is the following information.

**[0036]** For example, in a mass spectrum as illustrated in Fig. 2A, of which the horizontal axis represents a mass-to-charge ratio  $m/z$ , and the vertical axis represents a signal intensity, a plurality of peaks normally appear, including a peak originating from an intended target compound, peaks originating from impurities extraneous to a target compound, and noise peaks due to various factors other than the above. Such impurity-originated peak and noise peaks are desired to be absent on a mass spectrum, but such undesired peaks inevitably appear on a mass spectrum obtained through actual measurement. Therefore, when a database is created based on measured results, a mass spectrum including such undesired peaks will be registered.

**[0037]** In addition, a mass spectrum obtained through actual measurement includes a peak that can be an obstacle to identify the compound, while originating from a target compound. For example, in the case where a target compound is one of similar compounds having substantially the same main skeleton but differing in partial structure from one another, an ion peak originating from the main skeleton with a high signal intensity appears on a mass spectrum. However, the similar compounds cause such an ion peak to appear in common, which is not only useless to identify the target compound but also will trivialize the influence of a peak originating from a partial structure specific to the compound.

**[0038]** Thus, the LC-MS system in the present embodiment is configured to add, for each peak appearing on a mass spectrum, nonuse-indication information indicating that the peak is not used in database searches, that is, the peak is ignored, so as to prevent such peaks being unnecessary for identification or rather acting as an obstacle for identification from affecting a search result. Here, this nonuse-indication information is divided into two kinds: first nonuse-indication information and second nonuse-indication information, which will be described later.

**[0039]** The nonuse-indication information can be set by a user as appropriate for mass spectrum data 332 on any compound having already been registered in the compound database 33, or for mass spectrum data 332 on any compound to be newly registered in the compound database 33. For example, when the user performs a predetermined operation using the input unit 4, the attribute information setting section 35 displays a mass spectrum of a compound specified by the user on a screen of the display unit 5. When the user specifies on the displayed mass spectrum a peak not to be used for identification by a clicking operation or the like using the input unit 4, the attribute information setting section 35 recognizes the mass-to-charge ratio of the specified peak and writes information thereon as nonuse-indication information into predetermined storage area where attribute information 333 is stored in the compound data-

base 33.

**[0040]** Of course, if the nonuse-indication information set in such a manner is rewritten due to an operation error or out of spite, the reliability of analysis will be decreased, thus security measures are desirably taken so that a new setting or change of nonuse-indication information can be performed by only a person in charge given a higher privilege than that of common operators.

**[0041]** Next, a characteristic compound identification process performed in the LC-MS system in the present embodiment with nonuse-indication information set in the compound database 33 as previously described is described with reference to Fig. 2A, Fig. 2B, and Fig. 2C.

**[0042]** On a chromatogram created based on data collected through measurement performed on a target sample and displayed on the screen of the display unit 5, an operator specifies a peak estimated to originate from a compound intended to be identified, using the input unit 4. Then, data obtained at a time point of the peak top of the specified peak is read from the measurement data storage section 31, and the mass spectrum creating section 32 creates a mass spectrum based on this data and transmits the mass spectrum to the database search processing section 34. The database search processing section 34 searches the compound database 33 for a compound the spectral pattern of which is similar to this given mass spectrum (hereafter, referred to as a target mass spectrum), so as to identify an unknown compound intended.

**[0043]** The database search in the database search processing section 34 is performed in two steps: (I) rough extraction of compound candidates having similar spectral patterns, performed by the compound candidate extracting section 342, and (II) calculation of a score indicating similarity of spectral pattern in detail for each of the compound candidates, performed by the scoring section 343. Based on the calculated score, narrowing and ranking final candidates are performed by the candidate ranking section 344. The first nonuse-indication information previously described is referred to in both of the steps (I) and (II), and the second nonuse-indication information previously described is referred to in only the step (II).

**[0044]** In extracting the compounds having spectral patterns similar to the spectral pattern of the target mass spectrum, the compound candidate extracting section 342 reads mass spectrum data 332 corresponding to compounds registered in the compound database 33 one by one. At this point, the compound candidate extracting section 342 also reads the first nonuse-indication information stored in the form of the attribute information 333. The unnecessary information deleting section 341 deletes a peak to which first nonuse-indication information is set in the mass spectrum data, and the compound candidate extracting section 342 assesses whether or not the mass spectrum subjected to the peak deleting process is similar to the target mass spectrum in spectral pattern.

**[0045]** Now, assume that, of peaks on a mass spec-

trum illustrated in Fig. 2A collected in the compound database 33, flags of first nonuse-use information are set to peaks attached with the mark "O", and flags of second nonuse-use information are set to peaks attached with the mark "Δ". In this case, in the compound candidate extraction, the unnecessary information deleting section 341 deletes the peaks set with first nonuse-indication information, and thus the mass spectrum to be transmitted to the compound candidate extracting section 342 for the assessment of spectral pattern similarity becomes such as a mass spectrum illustrated in Fig. 2B. That is, the peaks set with the mark "O" in Fig. 2A are ignored in the assessment of spectral pattern similarity. Therefore, setting first nonuse-indication information to noise peaks accidentally appearing in actual measurement for the registration in the database enables compound candidate extraction being the same as in the case of using a mass spectrum in which such noise peaks are substantially absent.

**[0046]** A specific algorithm for the compound candidate extraction does not matter in particular, but the assessment of similarity may be performed in such a manner as to focus only on a mass-to-charge-ratio position at which a peak appears irrespective of the magnitude of the signal intensity of the peak as long as the signal intensity is not less than a predetermined threshold value, so as to search out compounds from the compound database 33 registering a huge number of compounds in a time as short as possible.

**[0047]** Next, for each of the plurality of extracted compound candidates, the scoring section 343 calculates a score obtained by quantifying the degree of similarity in spectral pattern between the target mass spectrum and a mass spectrum corresponding to the compound candidate. In this calculation of the score, both first nonuse-indication information and second nonuse-indication information stored in the compound database 33 are read as attribute information 333, and the unnecessary information deleting section 341 deletes, from read mass spectrum data, a peak with at least one of the first nonuse-indication information and the second nonuse-indication information set. Then, the scoring section 343 calculates a score based on, for example, differences in positions and signal intensities of peaks between the mass spectrum subjected to the peak deleting process and the target mass spectrum.

**[0048]** In the example illustrated in Fig. 2A, Fig. 2B, and Fig. 2C, since the unnecessary information deleting section 341 deletes, in scoring, the peak with at least one of the first nonuse-indication information and the second nonuse-indication information set, the mass spectrum to be transmitted to the scoring section 343 becomes such as a mass spectrum illustrated in Fig. 2C. That is, both of the peaks set with the mark "O" and the peaks set with the mark "Δ", in Fig. 2A, are ignored in scoring. Therefore, setting second nonuse-indication information to, for example, an ion peak originating from a main skeleton not useful for discriminating from similar compounds pre-

vents the similarity of such a peak from being reflected on the score and makes the similarity and the dissimilarity of another peak originating from a specific partial structure clearly appear in the score. As a result, the score of a compound candidate having a specific partial structure identical to the specific partial structure of the target compound is increased. In addition, setting first nonuse-indication information to a noise peak or the like as previously described prevents the noise peak from being reflected on the score, improving the accuracy of the score.

**[0049]** After calculating the score for each of the plurality of compound candidates, the candidate ranking section 344 exclude, if any, a compound candidate having an extremely low score from a list. Then, the candidate ranking section 344 sorts the remaining compound candidates in a descending order of scores, and displays the compound candidates on the screen of the display unit 5 together with their scores, as an identification result. As seen from the above, the LC-MS system in the present embodiment allows nonuse-indication information to be set to any peak on a mass spectrum registered in the compound database 33 and in searching the database, ignores some peaks using this nonuse-indication information to search for more likely compound candidates, so as to rank the compound candidates in terms of likelihood with high accuracy.

**[0050]** Next, a second embodiment of the LC-MS system using the mass spectrometry data-processing system according to the present invention is described with reference to the attached drawings. Fig. 3 is a schematic configuration diagram of the LC-MS system in this second embodiment, and Fig. 4 is a diagram illustrating an example of a compound identification process in the LC-MS system in the second embodiment. In Fig. 3, components identical to those of the configuration illustrated in Fig. 1 are denoted by the same reference numerals.

**[0051]** In this LC-MS system, attribute information 333 corresponding to mass spectrum data 332 stored in the compound database 33 includes priority indication information and intensity ratio threshold information. In addition, the database search processing section 34 includes a priority ion confirming section 345 and an intensity ratio computing section 346.

**[0052]** The priority indication information can be set, as with the nonuse-indication information in the first embodiment, by a user as appropriate for mass spectrum data 332 on any compound having already been registered in the compound database 33, or for mass spectrum data 332 on any compound to be newly registered in the compound database 33. Meanwhile, the intensity ratio threshold information is numerical value information that can be set, by a user as appropriate, to a peak on a mass spectrum set with priority indication information. This intensity ratio threshold information is for indicating a threshold for the assessment of a relative signal intensity of a peak intended (i.e., set with the priority indication information) to a predetermined reference intensity, which is described later. Specifically, for example, the

reference intensity may be determined in the form of the signal intensity of a peak at a specified mass-to-charge ratio on the mass spectrum or a maximum intensity observed on the mass spectrum.

**[0053]** Next, a characteristic compound identification process performed in the LC-MS system in the present embodiment with priority indication information and intensity ratio threshold information set in the compound database 33 as previously described is described with reference to Fig. 4.

**[0054]** As in the first embodiment, an operator specifies, on a chromatogram, a peak estimated to originate from a compound intended to be identified, using the input unit 4. Then, the database search processing section 34 searches the compound database 33 for a compound similar in target mass spectrum in spectral pattern. In this database search, the compound candidate extracting section 342 performs the rough extraction of compound candidates having similar spectral patterns, and subsequently, the scoring section 343 performs the calculation of a score indicating similarity of spectral pattern in detail for each of the compound candidates. Here, unlike the first embodiment, no nonuse-indication information is set as attribute information, and thus all peaks in mass spectrum data 332 registered in the compound database 33 are used in the database search.

**[0055]** It is assumed here, for ease of description, that the scoring section 343 performs the scoring based only on the degree of match between the mass-to-charge ratio of a peak on the target mass spectrum and the mass-to-charge ratio of a peak on the mass spectrum of each compound candidate, and the similarity between the peaks in signal intensity is ignored. As illustrated in Fig. 4, it is assumed that peaks are observed at three mass-to-charge ratios,  $m/z = a, b,$  and  $c,$  in the target mass spectrum. On the mass spectrum of each compound candidate, the scoring section 343 checks whether or not peaks are present at the three mass-to-charge ratios (actually, within mass-to-charge-ratio ranges having predetermined margins across the mass-to-charge ratios) and increments a score by one whenever a peak is present at the same mass-to-charge ratio as one of the three mass-to-charge ratios. Now, when the mass spectra of four compound candidates A, B, C, and D are such as those illustrated in Fig. 4, the calculation of scores based on matches between the mass-to-charge ratios of peaks yields a score of three for the compounds A and D because the number of peaks of matched mass-to-charge ratios is three, and a score of two for the compounds B and C because the number of peaks of matched mass-to-charge ratios is two.

**[0056]** In parallel to or following this process, the priority ion confirming section 345 checks whether or not priority indication information is set to a peak for which a score is incremented. Then, when priority indication information is set to a peak for which a score is incremented, the compound candidate is attached with a flag indicating the fact. In the example illustrated in Fig. 4, there

are two compounds A and C having priority indication information set to a peak for which a score is incremented. Thus, the two compound candidates are each attached with the flag. The candidate ranking section 344 excludes, if any, a compound candidate having an extremely low score from the list, sorts the remaining compound candidates in a descending order of scores, and displays the compound candidates on the screen of the display unit 5 together with their scores, as an identification result. At this point, a display indicating that a priority ion is contained is performed together on a compound candidate attached with a flag. In EXAMPLE 1 in Fig. 4, the text information indicating "PRIORITY ION PRESENT" is displayed, but the specific form of displaying is not limited to this.

**[0057]** In addition, there are cases where the presence/absence of a priority ion matters more when the difference between scores is small, and thus as illustrated in EXAMPLE 2 in Fig. 4, a compound containing a priority ion may be arranged at a higher rank even when the compound has a low score.

**[0058]** As seen from the above, in the result of the database search, a compound containing an ion set with priority indication information is discriminable from compounds not containing such an ion, and is displayed in some cases preferentially, that is, in a higher rank of a list, even having a low score. For this reason, for example, in the case where an ion characteristic of an intended compound or an ion useful for distinguishing the compound from the other compounds is known, setting priority indication information to such ions allows an operator to easily select a correct compound candidate from a search result.

**[0059]** In addition, in the case where intensity ratio threshold information is set together with priority indication information when the priority ion confirming section 345 attaches a flag indicating the presence of priority indication information to a compound candidate in accordance with the presence/absence of the setting of the priority indication information, the intensity ratio computing section 346 can perform the following process. That is, the intensity ratio computing section 346 determines, on the mass spectrum of the compound candidate, the signal intensity value of a peak considered to be an intensity reference of a relative intensity ratio and the signal intensity value of a peak set with priority indication information, to calculate an intensity ratio. Then, the intensity ratio is compared with a threshold indicated by the intensity ratio threshold information, and if the intensity ratio is equal to or greater than the threshold, the priority indication information is assessed to be effective. Then, a flag is attached to a compound candidate containing only an ion set with the priority indication information assessed to be effective. This allows a display indicating that a compound candidate contains a priority ion to be performed in a search result only when the compound candidate contains an ion set with the priority indication information, and at the same time when the signal inten-

sity of the ion reaches a reliable degree. In addition, it is allowed to perform a display indicating that a priority ion is contained irrespective of the result of assessing the intensity ratio, and moreover, to perform a display indicating the result of assessing whether the intensity ratio is equal to or greater than the threshold indicated by the intensity ratio threshold information.

**[0060]** As is clear from the previous description, the nonuse-indication information used in the LC-MS system in the first embodiment can be used concurrently with the priority indication information used in the LC-MS system in the second embodiment. That is, it may suffice to perform a database search excluding an ion peak set with nonuse-indication information to extract compound candidates, then assess whether or not an ion peak set with priority indication information is contained, and make the result of the assessment being reflected in the display of search result. This allows the exclusion of compounds other than a target compound and having a structure similar to the structure of the target compound, from a search result, and further, an explicit indication of compound candidates having a partial structure characteristic of the target compound, in the search result.

**[0061]** It should be noted that the previously-described embodiments are an application of the present invention to an LC-MS, and the present invention is applicable to general mass spectrometers performing compound identification using mass spectra registered in a compound database such as an LC-MS/MS, a GC-MS, a GC-MS/MS, as well as a stand-alone mass spectrometer not in combination with an LC or a GC.

**[0062]** It should be noted that the previously-described embodiments are a mere example of the present invention, and not only the previously-described modification but also any change, modification or addition appropriately made within the scope consistent with the present invention will evidently fall within the scope of claims of the present patent application.

#### REFERENCE SIGNS LIST

#### **[0063]**

1	Liquid Chromatograph	
2	Mass Spectrometry Section	
3	Data-Processing Section	
31	Measurement Data Storage Section	
32	Mass Spectrum Creating Section	
33	Compound Database	
331	Compound Information	
332	Mass Spectrum Data	
333	Attribute Information	
34	Database Search Processing Section	
341	Unnecessary Information Deleting Section	
342	Compound Candidate Extracting Section	
343	Scoring Section	
344	Candidate Ranking Section	
345	Priority Ion Confirming Section	

346	Intensity Ratio Computing Section
35	Attribute Information Setting Section

#### 5 Claims

1. A mass spectrometry data-processing system that estimates an unknown compound with a database search using a database in which mass spectra of known compounds are collected, the mass spectrometry data-processing system comprising:

a) a database in which mass spectra are collected, and nonuse-indication information indicating that it is not used in a step in a database search can be set as attribute information in association with at least some of ion peaks appearing on a mass spectrum; and

b) a search processing section for performing a database search through the use of a mass spectrum from which an ion peak with the nonuse-indication information attached in the attribute information is excluded when performing the database search.

2. The mass spectrometry data-processing system according to claim 1, wherein

the search processing section includes: a candidate extracting section for comparing patterns of mass spectra to extract a plurality of compounds to be candidates; and a ranking section for ranking the plurality of extracted candidates based on the similarities of mass spectra,

the nonuse-indication information includes: first nonuse-indication information indicating not being used in both of extracting the compounds by the candidate extracting section and ranking the compound candidates by the ranking section; and second nonuse-indication information indicating not being used only in ranking the compound candidates by the ranking section, and

the candidate extracting section extracts a plurality of compound candidates using mass spectra from which an ion peak attached with the first nonuse-indication information is excluded, and the ranking section ranks compound candidates using mass spectrum from which an ion peak attached with at least one of the first nonuse-indication information and the second nonuse-indication information is excluded.

3. A mass spectrometry data-processing system that estimates an unknown compound with a database search using a database in which mass spectra of known compounds are collected, the mass spectrometry data-processing system comprising:

a) a database in which mass spectra are collect-

ed, and priority indication information indicating that it is used preferentially in a database search can be set as attribute information in association with at least some of ion peaks appearing on a mass spectrum; and

5

b) a search processing section for performing a database search by comparing a mass spectrum of an unknown compound with mass spectra in the database to extract compound candidates that is supposed to be an unknown compound of interest, the search processing section extracting the compound candidates on one of conditions that an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound, or creating a search result that allows, of compound candidates extracted through a database search not using the priority indication information, a compound candidate of which an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound to be distinguished from other compound candidates.

10

15

20

4. The mass spectrometry data-processing system according to claim 3, wherein the database is configured to allow threshold information to be set, as attribute information, to an ion peak set with the priority indication information in addition to the priority indication information, the threshold information being used for assessing an intensity ratio of a signal intensity of the ion peak to a signal intensity of a specific peak on a mass spectrum in which the ion peak is present.

25

30

35

5. The mass spectrometry data-processing system according to claim 4, wherein the search processing section is configured to extract compound candidates on one of conditions that an ion peak set with the priority indication information is present on the mass spectrum of the unknown compound, and, on the mass spectrum of the unknown compound, the intensity ratio of the signal intensity of the ion peak to the signal intensity of the specific peak is equal to or greater than the threshold information set to the ion peak together with the priority indication information.

40

45

50

55

Fig. 1

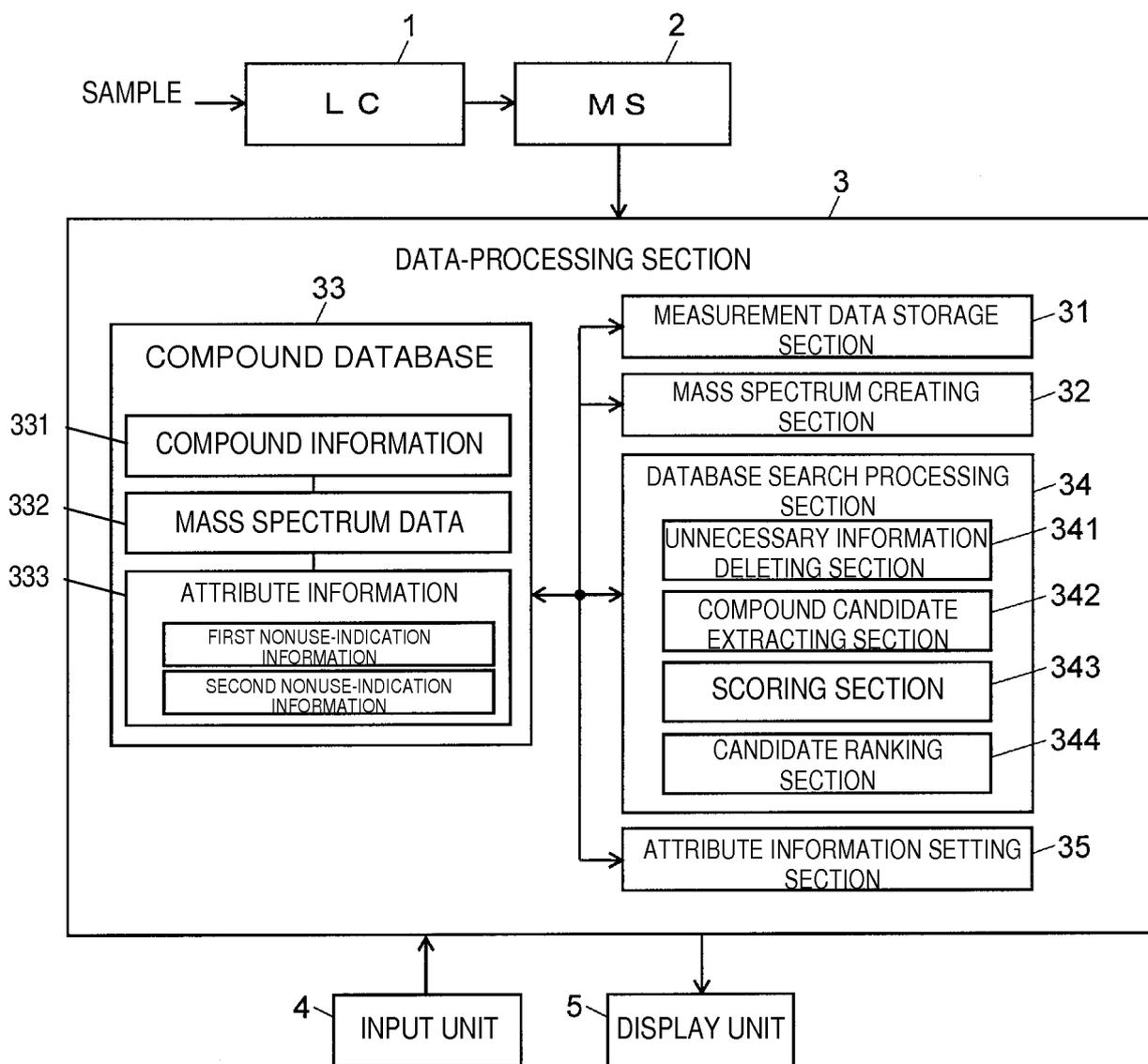


Fig. 2A

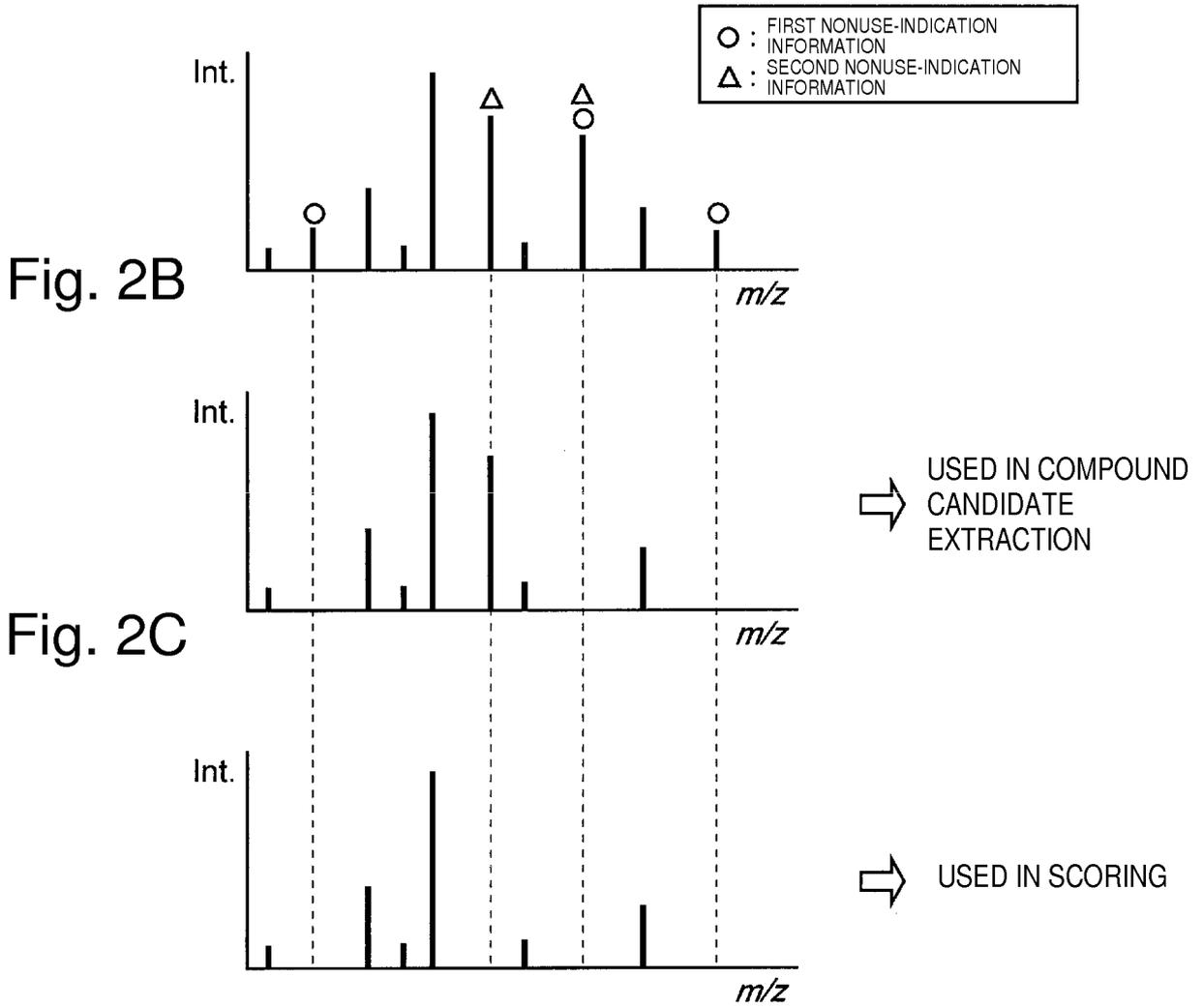


Fig. 3

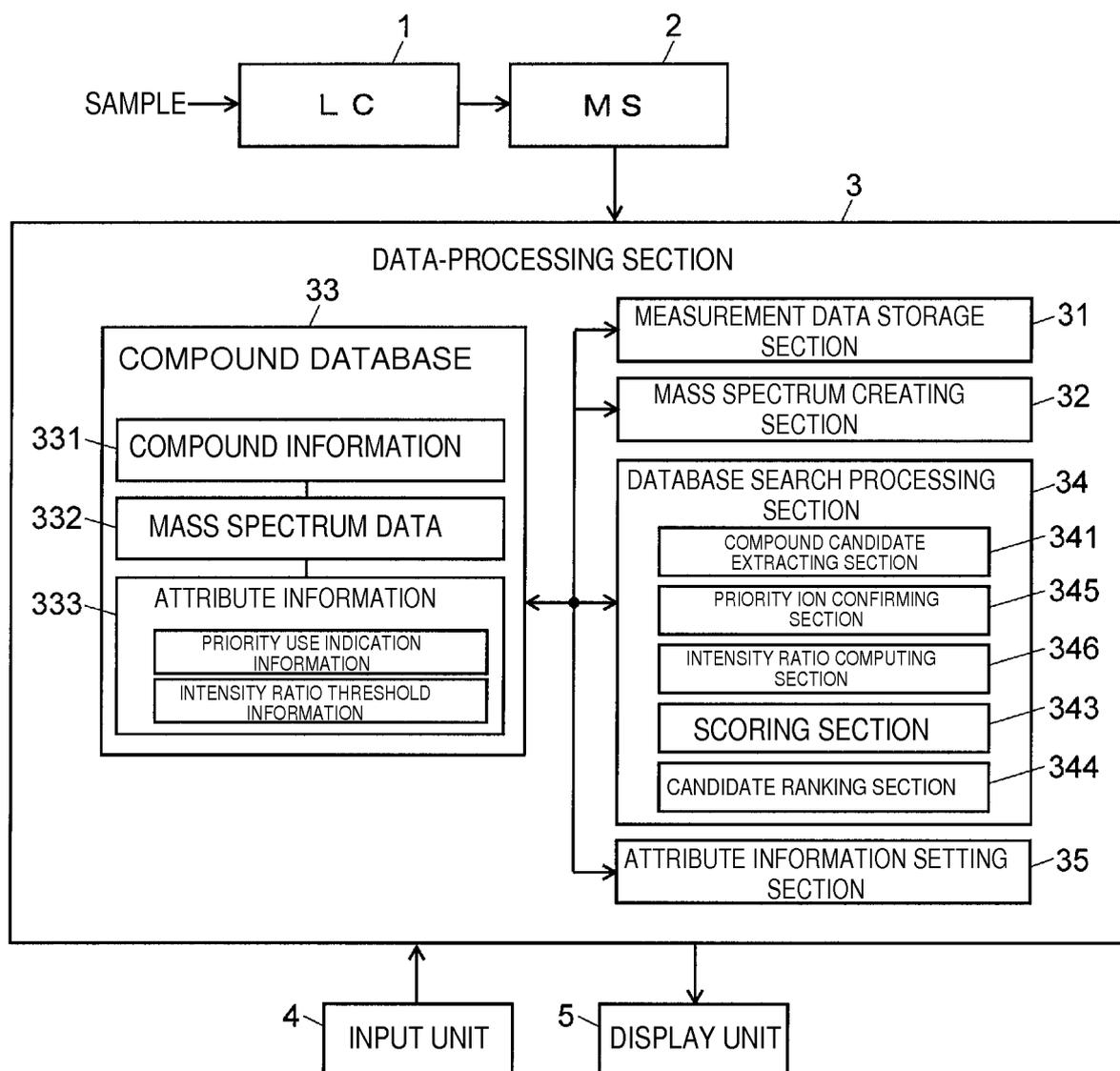
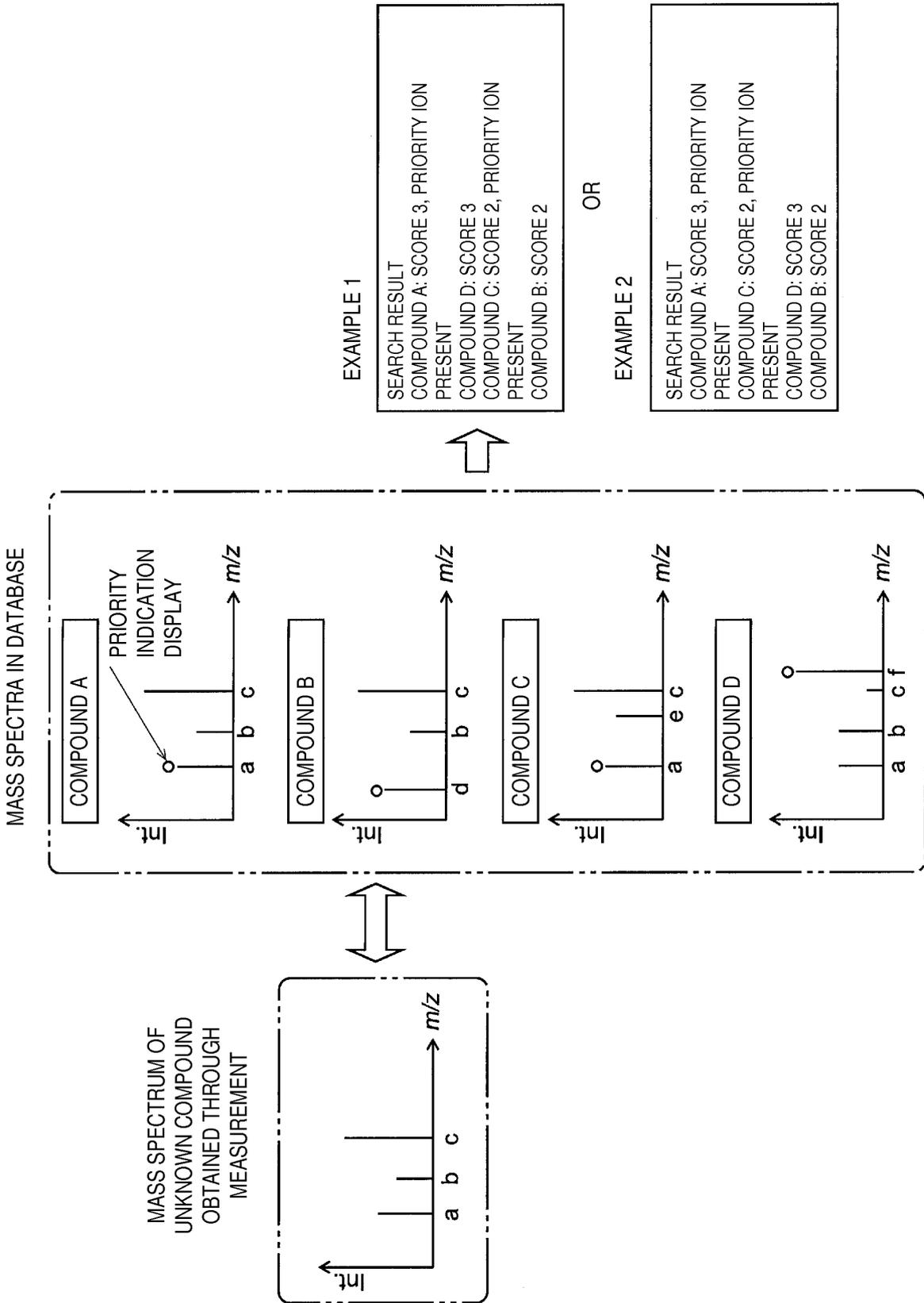


Fig. 4



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/050622

5	A. CLASSIFICATION OF SUBJECT MATTER G01N27/62(2006.01) i	
	According to International Patent Classification (IPC) or to both national classification and IPC	
10	B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G01N27/62	
15	Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2015 Kokai Jitsuyo Shinan Koho 1971-2015 Toroku Jitsuyo Shinan Koho 1994-2015	
20	Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JSTPlus/JMEDPlus/JST7580(JDreamIII), CiNii	
	C. DOCUMENTS CONSIDERED TO BE RELEVANT	
	Category*	Citation of document, with indication, where appropriate, of the relevant passages
25	X A	JP 2001-249114 A (Hitachi, Ltd.), 14 September 2001 (14.09.2001), paragraphs [0017] to [0038]; fig. 1 to 2, 4 to 7 & US 2001/0007349 A1
30	A	JP 2007-287531 A (Shimadzu Corp.), 01 November 2007 (01.11.2007), paragraphs [0025] to [0031]; fig. 2 to 4 & US 2008/0067344 A1
35	A	JP 5-149935 A (Shimadzu Corp.), 15 June 1993 (15.06.1993), paragraphs [0009] to [0015]; fig. 1 to 4 (Family: none)
40	<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.	
45	* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
50	Date of the actual completion of the international search 27 February 2015 (27.02.15)	Date of mailing of the international search report 10 March 2015 (10.03.15)
55	Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer  Telephone No.

Form PCT/ISA/210 (second sheet) (July 2009)

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/050622

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5939612 A (Hewlett-Packard Co.), 17 August 1999 (17.08.1999), column 3, line 62 to column 4, line 41; fig. 3 to 4 (Family: none)	1-5
A	Michiko UEBORI et al., "Study on Mass Spectral Database Obtained by Liquid Chromatography/Mass Spectrometry for Retrieval System of Chemicals", Journal of Japan Society for Analytical Chemistry, 05 March 2005 (05.03.2005), vol.54, no.3, pages 211 to 219	1-5
A	Takaaki NISHIOKA, "Taisha Busshitsu no Mass Spectrum Database", Annual Report of Institute of Natural Medicine University of Toyama, 31 March 2008 (31.03.2008), vol.34, pages 5 to 12	1-5

Form PCT/ISA/210 (continuation of second sheet) (July 2009)

**REFERENCES CITED IN THE DESCRIPTION**

*This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.*

**Patent documents cited in the description**

- JP 2005083952 A [0006]